

# **BCCWJ-TimeBank: Temporal and Event Information Annotation on Japanese Text**

**Masayuki ASAHARA\*, Sachi KATO\*, Hikari KONISHI\*,**

**Mizuho IMADA\*, and Kikuo MAEKAWA\***

## **Abstract**

Temporal information extraction can be divided into the following tasks: temporal expression extraction, time normalisation, and temporal ordering relation resolution. The first task is a subtask of a named entity and numeral expression extraction. The second task is often performed by rewriting systems. The third task consists of event anchoring. This paper proposes a Japanese temporal ordering annotation scheme that is used to annotate expressions by referring to ‘the ‘Balanced Corpus of Contemporary Written Japanese’ (BCCWJ). We extracted verbal and adjective event expressions as <EVENT> in a subset of BCCWJ and annotated a temporal ordering relation <TLINK> on the pairs of these event expressions and time expressions obtained from a previous study. The recognition of temporal ordering by language recipients tends to disagree with the normalisation of time expressions. Nevertheless, we should not strive for unique gold annotation data in such a situation. Rather, we should evaluate the degree of inter-annotator discrepancies among subjects in an experiment. This study analysed inter-annotator discrepancies across three annotators performing temporal ordering annotation. The results show that the annotators exhibit little agreement for time segment boundaries, whereas a high level of agreement is exhibited for the annotation of temporal relative ordering tendencies.

**Keywords:** Temporal Information Processing, Event Semantics, Corpus Annotation.

---

\*National Institute for Japanese Language and Linguistics, Japan  
E-mail: masayu-a@ninjal.ac.jp

## 1. Introduction

Temporal information processing in natural language texts has received increasing scholarly attention in recent years. Since the temporal orders of events often have implications for causal relations (cause and effect), identifying them is an essential task for deep understanding of language. Several types of resources for English temporal information processing have been developed, including an annotation specification TimeML (Pustejovsky *et al.*, 2003a) and annotated corpora, such as TimeBank (Pustejovsky *et al.*, 2003b) and Aquaint TimeML Corpus.

The English annotation specification has been extended as an International Standard Organization (ISO) standard of a temporal information mark-up language, namely ISO TimeML (Pustejovsky, Lee, Bunt, & Romary, 2010), which covers Italian, Spanish, Chinese, and other languages. Temporal information-annotated corpora in various languages have been developed and have been shared by natural language processing researchers. TempEval-2 (Verhagen, Sauri, Caselli, & Pustejovsky, 2010), a task for the SemEval-2010, and TempEval-3 (UzZaman *et al.*, 2013), a task for the SemEval-2013, have been proposed as shared temporal-relation reasoning tasks. In these shared tasks, datasets for English, Italian, Spanish, Chinese, and Korean are provided.

Nevertheless, there is no such resource for the Japanese language. In this paper, we present a means of porting a subset of ISO-TimeML into the Japanese language and describe the basic specifications of ‘BCCWJ-TimeBank,’ which is a realisation of the temporal information annotation of the ‘Balanced Corpus of Contemporary Written Japanese,’ or BCCWJ (Maekawa, 2008).

The purposes of temporal information annotation differ in accordance with the research goal. For natural language processing, users may need to develop and evaluate analysers for the annotation. Hence, for linguistic purposes, some users may want to formalise semantic expressions for temporal and event information expressions. Other users may want to evaluate human cognitive processes related to the expressions. The former purpose requires unique and consistent annotations. The latter purpose does not require consistent annotation; instead, it may require ways to evaluate variation among annotations. In this study, we conduct ‘pair annotation’ to formalise semantic expressions for temporal expressions. We also perform evaluation of the cognitive process of the expressions for temporal ordering annotation.

Porting TimeML into other languages can be challenging because of differences between languages. Even if we made a standardisation of ISO-TimeML, there would still be slight differences among the resources in terms of annotation targets, styles, formalism, philosophy, objectives, and focuses. Our research target is ‘temporal ordering’ in Japanese documents. We want to establish a machine learning-based temporal ordering analyser of the event and time

*Temporal and Event Information Annotation on Japanese Text*

expressions in Japanese. Before we develop the temporal ordering analyser, however, we have to analyse how well human annotators detect the temporal ordering. Therefore, our main research question in this article is to evaluate human annotators' skills in Japanese temporal ordering annotation. We permit inconsistency in semantic-level annotation among the annotators and quantitatively evaluate this inconsistency for the type of temporal ordering. Under the main research question, we did not perform sound and complete localization of ISO-TimeML to Japanese, as previous research has done.

The contributions of this work are as follows. First, to the best of our knowledge, this is the first corpus-based study on Japanese temporal information annotation. Second, we introduce two annotation paradigms for linguistic research. One paradigm is 'pair programming'-like annotation for consistent annotation. The other paradigm is annotation as a subjective experiment. Third, we evaluate cognitive processes in human temporal information processing.

The rest of this paper is organised as follows. Section 2 discusses previous studies related to this work. Section 3 briefly presents our annotation specification. Section 4 outlines the annotation processes of our work. Section 5 presents the corpus statistics and evaluations. Finally, Section 6 concludes this paper.

## **2. Previous Studies**

This section discusses previous studies on BCCWJ-TimeBank. Section 2.1 presents English temporal information processing. Section 2.2 presents ISO-TimeML, which is a standardisation of the annotation schema. Section 2.3 presents Asian language resources related to temporal information processing. Section 2.4 explains the target corpus of Japanese.

### **2.1 English Temporal Information Processing**

MUC-6 (Grishman & Sundheim, 1996) was a workshop on information extraction, which included temporal expression extraction as a subtask. TimeML <TIME3> tags used to be the *de facto* standard of normalization of temporal expressions; however, temporal information processing was then extended to event semantics. TimeML provides an annotation schema for event expressions and temporal relation extraction. Following this, TimeBank and some other corpora were developed. Using this corpus, machine learning-based temporal relation extraction methods have been developed (Boguraev & Ando, 2005; Mani, 2006). In addition, shared task workshops, including TempEval (Verhagen *et al.*, 2007), TempEval-2 (Verhagen *et al.*, 2010), and TempEval-3 (UzZaman *et al.*, 2013), have been held in more formalized evaluation settings.

## 2.2 ISO-TimeML: Standardisation of the Annotation Schema

The ISO Technical Committee (TC 37) has proposed several standards for language resources under the collective category ‘Terminology and other language and content resources’. The committee (SC) is divided into four areas. TC 37/SC 4 is charged with looking at annotation standards for all areas of natural language resources. This area includes six working groups (WG) to design language annotation specification mark-up languages, such as stand-off mark-up and XML. TC 37/SC 4/WG 2, the semantic annotation WG, discusses semantic annotation standards. The original TimeML developers and TC 37/SC 4/WG 2 defined ISO-TimeML as a Semantic Annotation Framework (SemAF)-Time, that is ISO-24617-1:2012, within the context of TC 37/SC 4.

TimeML and ISO-TimeML define four types of entities: <TIMEX3>, <EVENT>, <MAKEINSTANCE>, and <SIGNAL>. The <TIMEX3> tag specifies various attributes of time expressions, such as *tid*, *type*, *quant*, *freq*, *mod*, and *value*. The time expressions are categorized into four types: DATE, TIME, DURATION, and SET. The attribute *@value* includes the normalised values of the time expressions in a machine-readable format. The <EVENT> tag specifies various attributes of event expressions, including the class of the event, tense, grammatical aspect, polarity, and modal information. The <MAKEINSTANCE> tag presents the event instances expressed by <EVENT>-tagged expressions. Finally, the <SIGNAL> tag annotates elements to indicate how temporal objects are related amongst themselves.

TimeML and ISO-TimeML also define several types of links. Among these, <TLINK> expresses temporal order among instances of time expressions, event expressions, or both.

## 2.3 Time Information Annotation in Asian Languages

Japanese temporal information processing is still being developed. We have developed temporal expression extraction resources only as a subtask of named entity extraction. The IREX NE Task (Sekine & Isahara, 2000) includes time expressions as the target. Sekine, Sudo, and Nobata (2002) maintained an extended named entity hierarchy for Japanese and other languages, which includes five subcategories of time expressions and six subcategories of period expressions.

In the case of Chinese, Cheng developed a Time-ML compatible Chinese Temporal Annotation Corpus (Cheng *et al.*, 2008a) and proposed some models for the data (Cheng *et al.*, 2008b). This was the first localization work with TimeML for Chinese and was before ISO-TimeML. Cheng performed temporal ordering information annotation on 151 articles from the Penn Chinese Treebank (Xue *et al.*, 2005). In their work, syntactic dependency relations derived from the Penn Chinese Treebank were utilized for the annotation and temporal ordering estimation. There are two representative temporal information annotated

*Temporal and Event Information Annotation on Japanese Text*

corpora in addition to the abovementioned works: the Chinese part of the ACE 2005 multilingual training corpus (Walker *et al.*, 2006) and the TempEval-2 Chinese data sets (Xue & Zhou, 2010). The former is only for temporal expression extraction and normalization. The latter focuses on the temporal ordering of four relations, similar to this work, and it will be presented in Section 3.4. The TempEval-2 Chinese data sets are also based on 60 articles from the Penn Chinese Treebank that were analysed by a two-phase double blind and adjudication process. Nevertheless, sound and complete annotation cannot be achieved. Recently, the temporal expression annotation of TempEval-2 Chinese data was fixed by Li *et al.* (2014), and a Chinese temporal tagger based on the new annotation is publicly available.

In the case of Korean, KTimeML is a temporal information annotation guideline for Korean (Im *et al.*, 2009). Im and his colleagues utilized morpheme-based stand-off annotation and surface-based annotation.

A contrastive evaluation among Asian temporal information language resources is difficult. The research objectives vary among the previous articles. In addition, some detailed annotation guidelines are not in the previous articles but are written in manuals in their own language. Nevertheless, we emphasize that the temporal information processing of Asian languages is still an ongoing process. There are no sound and complete language resources on temporal information processing. One reason might be that the localization of all ISO-TimeML tags and attributes does not always help the temporal information processing. In addition, there are still language-independent issues for each language. Another reason might be that the human recognition system of temporal information is not stable among people. This article attempts to evaluate the human recognition system of temporal ordering in cognitive experiments.

## 2.4 BCCWJ and its Annotations

BCCWJ was publicly released in 2011 by the National Institute for Japanese Language and Linguistics (NINJAL) in Japan. It consists of three sub-corpora: ‘Publication,’ ‘Library,’ and ‘Special purpose’. ‘Publication’ consists of samples extracted randomly from books, magazines, and newspapers published during 2001-2005. ‘Library’ consists of randomly extracted samples from texts in circulation at libraries during the period 1986-2005. Finally, the ‘Special purpose’ sub-corpus consists of several mini-corpora without a statistical sampling method being used. It includes text from Yahoo! Answers, Yahoo! Blogs, white papers, and school textbooks. The total size of BCCWJ is about 100 million words.

The part of BCCWJ called ‘CORE’ manually annotates word boundaries, base phrase boundaries, and morphological information. CORE consists of six registers found in ‘Publication’ and ‘Special purpose’: books (PB), magazines (PM), and newspapers (PN) from ‘Publication,’ along with Yahoo! Answers (OC), Yahoo! Blogs (OY), and white papers

(OW) from ‘Special purpose’. The size of CORE is about 1.3 million words.

The BCCWJ data include several annotations, such as metadata, document structure, sentence boundaries, word boundaries, and phrase (bunsetsu) boundaries. NINJAL suggests two sorts of word delimitation definitions: one is a *short word unit* and the other is a *long word unit*. Each of these word delimitations is coded with the UniDic part-of-speech tag (Den et al., 2008).

Several research institutes have developed further linguistic annotations for CORE, such as syntactic dependency structures, developed by Nara Institute of Science and Technology (NAIST) and NINJAL; predicate-argument relations, developed by NAIST; named entities, developed by Tokyo Institute of Technology (TITECH); modality, developed by Tohoku and Yamanashi Universities; and Japanese framenet, developed by Keio University. The multi-word functional expressions are maintained by Tsukuba University. The CORE samples are split into annotation priority sets from A to E to allow the annotations to overlap as much as possible. Some of these annotations can be used as presupposed information for our annotation. In this way, BCCWJ is the most promising resource making linguistic annotations for both NLP and linguistic researchers.

Table 1 shows the basic statistics and priority sets of BCCWJ CORE. The word unit is based on the *short word unit*, UniDic standard (Den, Nakamura, Ogiso, & Ogura, 2008); UniDic is a lexicon for Japanese morphological analysis.

**Table 1. BCCWJ CORE: Registers and priority sets.**

Register	(Abbr.)	Priority Set	# of Samples	# of Words
White Paper	OW	A to D	62	197,011
Books	PB	A to D	83	204,050
Newspapers	PN	A to E	340	308,504
Yahoo! Answers	OC	A to B	938	93,932
Magazines	PM	A to D	86	202,268
Yahoo! Blog	OY	A to B	471	92,746

### 3. Temporal Information Annotation Specification

This section presents a specification for Japanese temporal information annotation. The annotation is realised as BCCWJ-TimeBank. The specification is based on TimeML (Pustejovsky et al., 2003a) and is adapted for the Japanese language. Figure 1 shows an example of the annotation. Below, we present an overview of the specification of TimeML tags: <TIMEX3> for temporal expressions, <EVENT> and <MAKEINSTANCE> for event expressions, and <TLINK> for temporal ordering. We mention other tags that we exclude

**PN23\_0001 Sample in BCCWJ CORE**

```

<TIMEX3 @value="2002-04-11" @definite="true" @tid="t0" functionInDocument="CREATION_TIME"
type="DATE"/><sentence>地方自治体が<EVENT @class="NULL" @eid="e25">運営する</EVENT>公営地下鉄二十六
路線のうち<TIMEX3 @value="FY2000" @definite="FALSE" @valueFromSurface="FY2000" @tid="t4"
@type="DATE">二〇〇〇年度</TIMEX3>決算で経常損益が黒字なのは、札幌市南北線など四路線に<EVENT
@class="I_ACTION" @eid="e26">とどまった</EVENT>ことが、公共交通事業協会が<TIMEX3 @value="2002-04-10"
@definite="true" @valueFromSurface="XXXX-XX-10" @tid="t5"
type="DATE">十日</TIMEX3><EVENT @class="I_ACTION" @eid="e27">まとめた</EVENT>報告書で<EVENT
@class="I_STATE" @eid="e28">分かった</EVENT>。</sentence>
<MAKEINSTANCE @eventID="e26" @eiid="ei26"/>
<MAKEINSTANCE @eventID="e27" @eiid="ei27"/>
<MAKEINSTANCE @eventID="e28" @eiid="ei28"/>
<TLINK @relTypeA="after" @relTypeB="after" @relTypeC="during" @task="DCT"
@timeID="t0" relatedToEventInstance="ei26"/>
<TLINK @relTypeA="after" @relTypeB="after" @relTypeC="after" @task="DCT"
@timeID="t0" @relatedToEventInstance="ei27"/>
<TLINK @relTypeA="after" @relTypeB="after" @relTypeC="after" @task="DCT"
@timeID="t0" @relatedToEventInstance="ei28"/>
<TLINK @relTypeA="vague" @relTypeB="equal" @relTypeC="during" @task="T2E"
@timeID="t4" @relatedToEventInstance="ei26"/>
<TLINK @relTypeA="vague" @relTypeB="before" @relTypeC="before" @task="T2E"
@timeID="t4" @relatedToEventInstance="ei27"/>
<TLINK @relTypeA="vague" @relTypeB="before" @relTypeC="before" @task="T2E"
@timeID="t4" @relatedToEventInstance="ei28"/>
<TLINK @relTypeA="after" @relTypeB="before" @relTypeC="during" @task="T2E"
@timeID="t5" @relatedToEventInstance="ei26"/>
<TLINK @relTypeA="contains" @relTypeB="after" @relTypeC="finishes" @task="T2E"
@timeID="t5" @relatedToEventInstance="ei27"/>
<TLINK @relTypeA="contains" @relTypeB="equal" @relTypeC="before" @task="T2E"
@timeID="t5" @relatedToEventInstance="ei28"/>
<TLINK @relTypeA="vague" @relTypeB="before" @relTypeC="contains" @task="E2E"
eventInstanceID="ei26" @relatedToEventInstance="ei27"/>
<TLINK @relTypeA="before" @relTypeB="before" @relTypeC="before" @task="E2E"
eventInstanceID="ei27" @relatedToEventInstance="ei28"/>
English Translation:
Municipal Transportation Works Association published a report on April 10th. The report shows that
only 4 public tube railways (e.g. Sapporo City Nanboku line) from 26 have a surplus.

```

**Figure 1. An example of Japanese BCCWJ TimeBank annotation.**

from the Japanese temporal information annotation. Then, we perform a comparison with the previous research on Asian languages.

**3.1 Overview of Japanese Temporal Information Annotation**

This section presents an overview of the Japanese temporal information annotation.

We introduce three types of tags for the annotation: i) the time expression tag <TIMEX3>, ii) the event expression and event instance tags <EVENT> and <MAKEINSTANCE>, iii) the temporal relation tag <TLINK>. We explain the specifications of each tag in the following subsections.

Note that our research objective is not the localization of ISO-TimeML for Japanese but is to investigate the ability of the human recognition process of temporal ordering. We omitted detailed localization of ISO-TimeML tags and attributes. In the original TimeML, <SIGNAL>, <SLINK>, and <ALINK> are defined. <SIGNAL> is used with some temporal prepositions and conjunctions in English, <SLINK> is used for subordination relations, and <ALINK> is used for non-constituent aspectual relations. Currently, we are not using these with the BCCWJ-TimeBank. We leave these directions for our future work for the following reason. According to the specification <SIGNAL> in TimeML, ISO-TimeML, and (Setzer, 2001), we have to annotate nearly all of the functional words (auxiliary verbs and postpositions) that should be tagged <SIGNAL>. If we do not introduce any subcategories for each functional word, annotating all of the functional words is the same as annotating none of the functional words. <SLINK> should be annotated after the annotation of the subordinate clause boundaries with the clause functions. Although <ALINK> should be annotated as an aspectual compound main verb structure in Japanese, verbs other than the main verb in a compound verb are defined as auxiliary verbs, according to the definition of the UniDic POS tagset. As presented in Section 2.3, the previous work on Chinese also focuses on <TLINK>.

### 3.2 <TIMEX3>: Temporal Expression Annotation

The target temporal expressions for <TIMEX3> are DATE, TIME, DURATION, and SET by @type. We do not permit any nests of <TIMEX3>. We clip the expressions according to characters because Japanese does not have word delimitation spaces.

The attributes of @tid, @type, @value, @freq, @quant, and @mod have been inherited from the original TimeML. There is an issue regarding which calendar to use in porting TimeML to Japanese. In Japan, we use not only the Western calendar but also a native Japanese calendar that is based on the year of the Emperor’s reign. We introduce a new attribute @valueFromSurface to address this issue. @valueFromSurface includes a @value-like string to indicate a machine-readable date/time value. @value includes the normalised version of value, whereas @valueFromSurface includes the non-normalised version of the value, which can be generated on rewrite rules. @valueFromSurface can encode Japanese calendars. For example, ‘平成 26 年’ (the 26th year of the Heisei era) is encoded in the @valueFromSurface as ‘H26’ and normalised as the @value of ‘2014’ in the ISO-8601-like format.

The difference between @value and @valueFromSurface shows the use of the normalisation procedure. Nevertheless, we cannot judge whether the <TIMEX3> is fully normalized (fully specified) or underspecified. We introduce another new attribute, @definite, to indicate whether the <TIMEX3> is fully specified ‘true’ or underspecified ‘false’.



### 3.3 <EVENT>, <MAKEINSTANCE>: Event Expression Annotation

Next, we need to annotate the event expressions and instances to link the temporal ordering to <TIME3>. The event expression candidates are automatically extracted from the BCCWJ of morphological information. We define *long word units* with verbs and adjectives, resulting in a total of 4,953 event expression candidates. First, the candidates are judged by two annotators as to whether the target expression is an event expression. If the expression boundaries are not valid, a longer expression covering the candidate is redefined by the annotators. Second, the annotators judge whether there are any instances of the target expression on the timeline. If an instance is recognised, the annotators define a <MAKEINSTANCE> in the corpus. The <MAKEINSTANCE> is a stand-off from the event expression, but is linked to the <EVENT> tag by the @eid attribute. Third, the annotators annotate the @class attribute on the <EVENT>. There are nine @class attributes: seven for event instances (OCCURRENCE, REPORTING, PERCEPTION, ASPECTUAL, I\_ACTION, I\_STATE, STATE) and two for non-instances (NULL and NONE). The difference between NULL and NONE is that the former is applied by <EVENT> annotators and the latter by <TLINK> annotators. The instances are double-checked by both <EVENT> and <TLINK> annotators. These attributes are described in more detail below.

- OCCURRENCE: These are event expressions without event arguments describing something that happens or occurs in the world (the argument event). Most event expressions belong to this class.
- REPORTING: These are event expressions with an event argument describing the action of an animate actor declaring, narrating, or informing about the argument event.
- PERCEPTION: These are event expressions with an event argument describing the physical perception of the argument event.
- ASPECTUAL: These are event expressions with an event argument describing some aspectual feature of the argument event.
- I\_ACTION: These are intentional action expressions with an event argument describing an action or situation to introduce the argument event, from which we can infer something, given its relation with the I\_ACTION.
- I\_STATE: These are intentional state expressions with an event argument referring to an alternative or possible world.
- STATE: These are state expressions in the timeline. We only annotate these when an instance is introduced and becomes an argument of other event expressions.
- NULL, NONE: These are non-instance expressions

The annotator discriminates whether the target is an event or a state (STATE). Then, he or she judges whether the target has an event argument (OCCURRENCE). Finally, he or she

categorises any target with an event argument into one of the five categories of REPORTING, PERCEPTION, ASPECTUAL, I\_ACTION, and I\_STATE.

The two annotators and two supervisors defined a detailed linguistic annotation specification employing some Japanese language tests that are based on linguistic research (Kudo, 1995, 2004; Nakamura, 2001). The two annotators were trained on the specification until the agreement rate reached 75%.

<MAKEINSTANCE> attributes such as @tense, @aspect, and @modality are not well-maintained in the current status. Japanese tense and aspect from the surface forms cannot be matched to ISO-TimeML. Japanese ‘surface tense’ is marked by *-u* and *-ta* and ‘surface aspect’ is marked by *-u* and *-teiru*.

In the case of tense, the surface tense only expresses the difference between past and non-past. Furthermore, the event marked by the surface tense may have discrepancies with the temporal ordering on the timeline. Nevertheless, the temporal ordering relation <TLINK> between DCT and the target event expresses ‘deep tense’ information. The deep tense information is a translingual feature. When we want to investigate the discrepancy between surface tense and deep tense, it can be done by comparing tense morphemes from the *short word unit* and <TLINK> with DCT.

In the case of aspect, Japanese has some alternative surface aspectual expressions such as *-tearu* and *-teoru*. These aspectual expressions, excluding *-ru* and *-teiru*, are called ‘semi-aspectual expressions’ in Kudo (1995). These rich surface aspectual systems cannot be mapped on the ISO-TimeML original labels. Furthermore, the aspectual systems in Japanese vary among regions or dialects (Kudo, 2004). We believe this issue is beyond the scope of our current work, and we will treat this issue in our future work.

In the case of modality, Yamanashi University and Tohoku University developed annotation of modality information in BCCWJ. The researchers focused on whether the target event happened or did not happen in the real world in their annotation schema. Their research objective is different from ours. Nevertheless, they provide rich information on tense aspect and modality structure in their research. Layering their annotation on BCCWJ-TimeBank makes the corpus more informative.

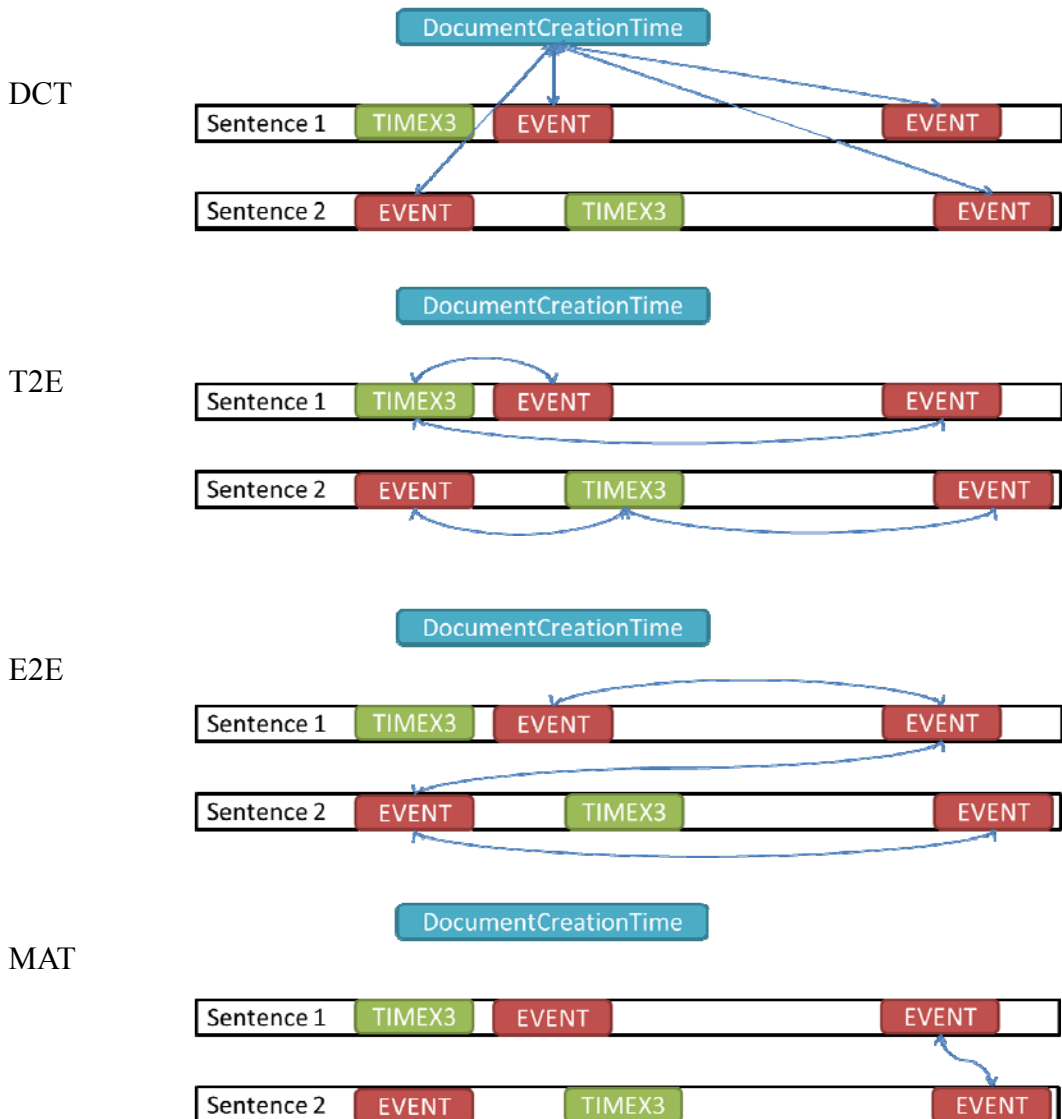
### 3.4 <TLINK>: Temporal Ordering Annotation

<TLINK> defines the temporal ordering of temporal information expressions and event expressions. We used a variant of Allen’s interval algebra as <TLINK> labels; there are 13 labels for temporal ordering and three for event-subevent relations.

We also have one label ‘vague’ for underspecified relations. Figure 3 shows the thirteen labels for temporal ordering and the three for event-subevent relations. The three underlined

## Temporal and Event Information Annotation on Japanese Text

labels, namely ‘is\_included,’ ‘identity,’ and ‘includes,’ are event-subevent versions of ‘during,’ ‘equal,’ and ‘contains,’ respectively. Strictly speaking, we can also define event-subevent versions of ‘finishes,’ ‘started-by,’ ‘starts,’ and ‘finished-by’. We did not define these, however, because they are rare and TimeML did not define them.



Note that Japanese is a strictly head-final language.

The matrix verb phrases tend to be put near the end of sentence.

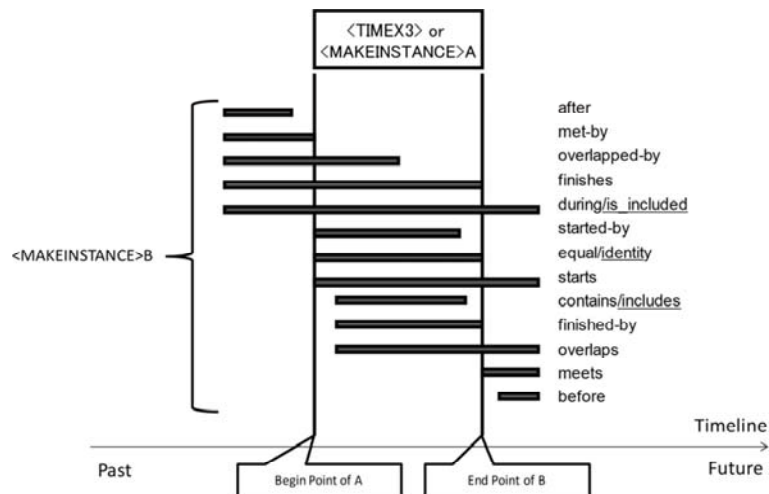
*Figure 2. <TLINK>: The four types of relations.*

<TLINK> annotators are different from <EVENT> and <MAKEINSTANCE> annotators. Three annotators annotate the <TLINK> labels on some of the pairs among <TIMEX3> and <MAKEINSTANCE>. The number of <TLINK> candidates is equal to the combination number of <TIMEX3> and <MAKEINSTANCE>. It is difficult to check all possible pairs in the documents; therefore, we limited the target pairs to the following four types of relations:

- ‘DCT’: relations between a <TIMEX3> of document creation time (DCT) and an event instance;
- ‘T2E’: relations between a <TIMEX3> (non DCT) and an event instance within one sentence;
- ‘E2E’: relations between two consecutive event instances; and
- ‘MAT’: relations between two consecutive matrix verbs of event instances.

Figure 2 presents the four types of relations.

If the relation is between two different possible worlds, we use the label ‘vague’. When we regard the ‘vague’ relations as disjointed links, the connected subgraph indicates the different possible worlds. The value of <TIMEX3> is regarded not as a time point but as a time interval. The event instance of a punctual verb is regarded as a time point occurrence, whereas the other event instances are regarded as time interval occurrences. Figure 3 presents an overview of the <TLINK> labels.



**Figure 3.** <TLINK> labels.

### 3.5 Other Tags in TimeML and ISO-TimeML

We will annotate <SIGNAL>, <SLINK>, and <ALINK> in the original TimeML tag in the future. In terms of the BCCWJ annotation background, several institutes are working together to take responsibility for layers of annotations. Some tags in TimeML originals presuppose other layers' annotation. To reduce annotation cost and to keep consistency among the professionals in other institutes, we postpone <SIGNAL>, <SLINK>, and <ALINK> until the presupposed annotation is finished. For example, <SIGNAL> tags are highly related to the multi-word functional words. <SLINK> tags are related to the subordinate clause structures in the treebank annotation. <ALINK> tags are related to the compound verb construction annotation.

### 3.6 Comparison with other Temporal Information Annotated Corpora in Asian Languages

This section presents a contrastive comparison with the previous research on Asian language resources.

First, the IREX NE Task data in Japanese (Sekine & Isahara, 2000) and the Chinese part of the ACE 2005 multilingual training corpus (Walker *et al.*, 2006) focused only on time expression extraction and normalization.

Second, both of Cheng's works (Cheng *et al.*, 2008a) and the TempEval-2 Chinese data sets focus on automatic <TLINK> annotation. <SLINK> and <ALINK> are not annotated on these resources. <TLINK> in Cheng's work is an annotated subset of the event and time expression pairs, and it expands the relation among the transition rules of temporal logic. On the other hand, <TLINK> in the TempEval-2 Chinese data sets targets limited pairs of the event and time expressions, which is almost the same as our approach, described in Section 3.4.

Third, KTimeML (Im *et al.*, 2009) is a TimeML compatible annotation schema. Nevertheless, the researchers introduce several changes to the original TimeML markup philosophy, including a change from word-based in-line annotation to morpheme-based stand-off annotation. In the annotation definition, we inherit some <TIMEX3> standards from KTimeML, such as introduction of the week of the month.

Fourth, we discuss the word segmentation issue in Asian languages. CJK languages are written without word boundaries. The two Chinese resources are based on the Penn Chinese Treebank, and they use the word unit from the original Penn Chinese Treebank. KTimeML uses a morpheme unit from Korean. Nevertheless, they introduce the stand-off style annotation over the morpheme. BCCWJ-TimeBank uses the *short word unit* of the original BCCWJ.

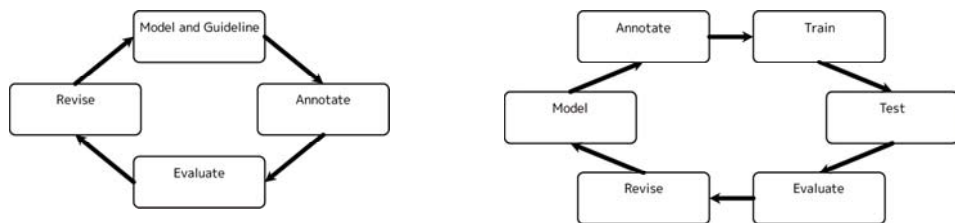
There are still differences in several layers of the annotation among the Asian temporal information annotated language resources. Space limitations do not permit a full discussion of these differences.

## 4. Annotation Processes

This section presents the annotation processes. First, we present the MAMA-cycle and MATTER-cycle. Then, we introduce two additional paradigms for the annotation processes: MAMA-cycle with a pair-programming type of method for <TIMEX3> and annotation as a cognitive science experiment for <TLINK>. Note that we cannot introduce new annotation methods for <EVENT> and <MAKEINSTANCE>. We performed these two annotations by a simple MAMA-cycle.

### 4.1 MAMA-cycle and MATTER-cycle

O'Reilly's book, *Natural Language Annotation for Machine Learning* (Pustejovsky & Stubs, 2012), presents two types of annotation cycles. The MAMA-cycle, whose initials represent Model-Annotate-Model-Annotate, stresses the importance of creating a guideline or specification. The MATTER-cycle, whose initials represent Model-Annotate-Train-Test-Evaluate-Revise, stresses the importance of creating the language analysers. Train and Test are the phases of machine learning in creating language analysers. Figure 4 presents the two types of cycles.



**Figure 4. MAMA-cycle and MATTER-cycle.**

The two cycles are aimed at producing effective guidelines, good language analysers, or both. Nevertheless, our research objective is different. The BCCWJ-TimeBank guidelines inherit most of the original TimeML/TimeBank schema. Our contribution to these guidelines is limited to the localization of the schema. We may produce language analysers in future work; however, we did not aim at the development of a temporal ordering relation analyser in the current stage. Our research objective is somewhat related to cognitive science because we would like to evaluate the human annotators' cognitive process of temporal ordering. Thus, in this work, we propose two additional annotation paradigms: MAMA-cycles with pair annotation and Annotation as a cognitive science experiment.

## 4.2 MAMA-cycle with Pair Annotation for <TIMEX3>

We used XML Editor oXygen 3 for <TIMEX3> annotation, and we defined DTD for BCCWJ-TimeBank. The DTD enables us to use the machine-aided (such as XML validation and a completion mechanism) environment of oXygen. An annotator performs in-line annotation on the original text corpus. We introduce a pair-programming type of method in which a display is shared by an annotator and supervisor. Although the method is stressful for both the annotator and supervisor, the data become more consistent and annotation errors are reduced. Figure 5 illustrates the proposed annotation process.

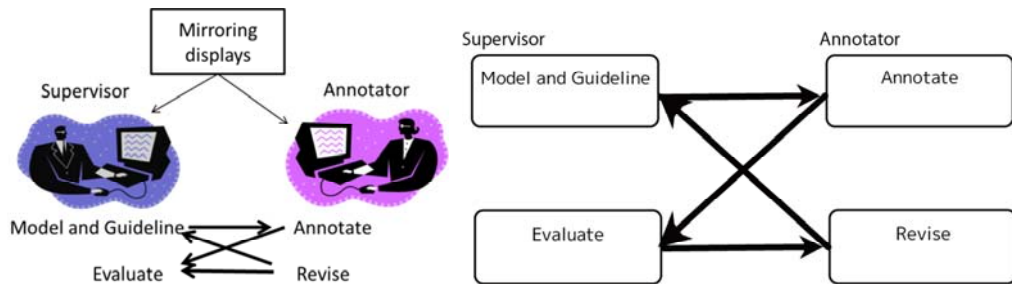


Figure 5. MAMA-cycle with a pair-programming method.

## 4.3 Annotation as a Cognitive Science Experiments for <TLINK>

Next, we performed a cognitive science experiment for <TLINK> annotations. In this paradigm, we evaluated the human cognitive process for perceiving the timeline. Here, we explain the process in detail. First, the supervisor gives annotation guidelines to three annotators. Second, the three annotators individually annotate <TLINK> information on the same dataset. Finally, the researchers evaluate the variance and differences among the three annotations. During this process, the annotators perform individual, not inter-annotator, revision. Figure 6 illustrates the proposed annotation process.

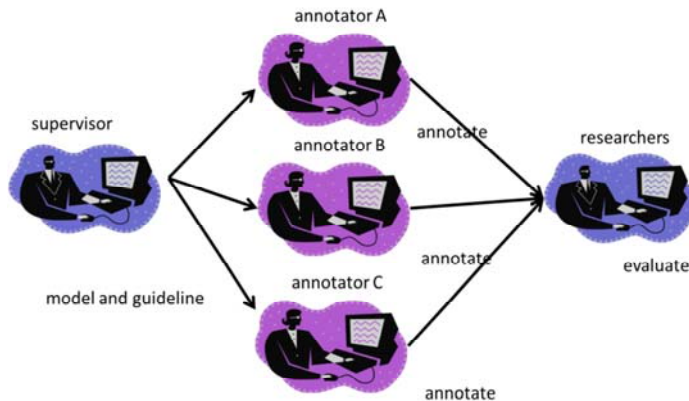


Figure 6. Annotation as a cognitive science experiment.

## 5. Corpus Statistics and Evaluations

This section presents basic statistics on BCCWJ-TimeBank, the Japanese corpus annotated for temporal information. We also consider the annotation environment of BCCWJ-TimeBank.

### 5.1 <TIMEX3>

Table 2 shows annotation target samples for <TIMEX3>. The column ‘W/TIMEX’ shows the number of samples or sentences that include at least one temporal information expression. Some samples in the registers OW (white paper), OC (Yahoo! Answers), and OY (Yahoo! Blogs) do not include any temporal information expressions.

**Table 2. <TIMEX3>: Annotation target samples.**

Register (Priority set)	# of Samples			# of Sentences			# of Words
	ALL	w/TIMEX	(Rate)	ALL	w/TIMEX		ALL
OW (A)	17	16	(94%)	1439	405	(28%)	58336
PB (A)	25	25	(100%)	2568	289	(11%)	57929
PN (A, B)	110	110	(100%)	5582	1562	(28%)	116834
OC (A)	518	250	(48%)	3479	488	(14%)	60086
PM (A)	23	23	(100%)	3066	413	(13%)	59372
OY (A)	257	198	(77%)	3986	765	(19%)	63459

Table 3 shows the basic statistics of <TIMEX3> annotations. The table shows the number of <TIMEX3> by @type and @definite and shows the relation of {@value and @valueFromSurface}. @type has four labels: DATE, TIME, DURATION, and SET. We exclude document creation time (DCT), which is given in corpus metadata, from the statistics. Then, we analyse the statistics on the basis of two perspectives. The first is whether @definite is ‘true’ or ‘false,’ in other words, whether the temporal information expression is fully specified or underspecified. The fully specified expression can be mapped on the timeline, while the underspecified expression cannot. The second perspective is whether @value and @valueFromSurface are identical (‘=’) or not (‘≠’). The former have undergone some normalisation procedure from the annotators, while the latter have not.

A total of 5,297 temporal information expressions were annotated in the corpus. Of those, 1,639 (30%) are fully specified expressions without any normalisation procedures applied. Further, 2,023 (37%) of the total can be normalised by contextual information, and 1,875 (34%) cannot. The third group needs more external information to be able to be normalised.



## Temporal and Event Information Annotation on Japanese Text

**Table 3. <TIMEX3>: @type×@definite×{@value, @valueFromSurface}.**

@definite	True (fully specified)			False (underspecified)		
@value and @valueFromSurface	All	=	≠	All	=	≠
DATE	2214 (61%)	381 (10%)	1833 (50%)	1438 (39%)	1275 (35%)	163 (4%)
TIME	188 (37%)	1 (0%)	187 (37%)	315 (63%)	239 (48%)	76 (15%)
DURATION	1129 (92%)	1128 (92%)	1 (0%)	99 (8%)	99 (8%)	0
SET	131 (85%)	129 (84%)	2 (2%)	23 (15%)	22 (14%)	1 (1%)
ALL	3662 (66%)	1639 (30%)	2023 (37%)	1875 (34%)	1635 (30%)	240 (4%)

Count (Rate)

In the ‘DATE’ expressions, most of the fully specified expressions (@definite ‘true’; 61%) have had manual normalisation performed (@value ≠ @valueFromSurface; 50%). This fact shows that the normalisation procedure is important for temporal information processing. The normalisation includes conversion from the Japanese to the Western calendar, including, conversion from a 2-digit to a 4-digit calendar, and completion year (taken from the document creation time).

In the ‘TIME’ expressions, most fully specified expressions have had manual normalisation performed. The normalisation includes completion date (from the document creation time) and resolution of a.m./p.m. ambiguity.

In the ‘DURATION’ and ‘SET’ expressions, @definite ‘true’ means that the length of the temporal region on the timeline can be uniquely determined. When we map this on the timeline, we need <TLINK> information with ‘DATE’ or ‘TIME’ expressions or event expressions.

Note that we reduce the annotation target samples of <EVENT>, <MAKEINSTANCE>, and <TLINK> PN register (A), which total 54 samples. The reason is that only the PN (newspaper) sample has date-level document creation time information as metadata. Table 4 shows the statistics of <TIMEX3> in the PN (A) samples.

**Table 4. <TIMEX3>: Statistics in PN (A).**

DCT or class	Count
DCT(DATE)	54
DATE	727
TIME	107
DURATION	291
SET	19
ALL	1198

## 5.2 <EVENT> and <MAKEINSTANCE>

We annotate <EVENT> and <MAKEINSTANCE> tags only on the PN register (A). Table 5 shows the statistics of <EVENT> tags by @class. Event instances by <MAKEINSTANCE> are defined on the last seven @class in the tables. The number of <MAKEINSTANCE> is 3,839.

**Table 5. <EVENT>: Statistics in PN (A).**

<EVENT>@class	count
Non-instance	(1129)
NULL	1114
NONE	15
Event instance w/o event arg	(2352)
OCCURRENCE	2352
Event instance w/ event arg	(1291)
REPORTING	126
PERCEPTION	27
ASPECTUAL	63
I_ACTION	880
I_STATE	195
State instance	(181)
STATE	181

## 5.3 <TLINK>

The three annotators were independently trained for <TLINK> annotation. The annotation was performed on four types of relations: ‘DCT,’ ‘T2E,’ ‘E2E,’ and ‘MATRIX’.

Table 6 shows annotation agreement among the 13+3+1 labels by the three annotations and relation types. The three  $\cap$ -connected numbers are the label counts by each of the three annotators. The right number after ‘=’ is the agreed count.

The agreed counts for ‘after,’ ‘during,’ ‘contains,’ and ‘before’ are higher than the others. These relations do not exhibit boundary matching between the two time intervals. The relation ‘equal’ is the most frequent of those that do include interval boundary matching. Other relations are infrequent and show low agreement counts

across the three annotators. These findings reveal that a judgment of interval boundary matching is rare and is difficult for human annotators. The relation ‘vague’ was agreed on 314 times by the three annotators. This fact shows that the discrimination of possible worlds might

## Temporal and Event Information Annotation on Japanese Text

be possible by annotation.

**Table 6. <TLINK>: Annotation agreement by Annotators × Labels × Relation types.**

Relation types	DCT	T2E	E2E	MATRIX	All
Count	3839	2188	2972	1245	10244
after	2352∩2326∩2133=1961	396∩441∩432=315	627∩631∩639=432	292∩284∩277=198	3667∩3682∩3481=2906
met-by	0∩0∩0=0	5∩10∩2=2	18∩12∩3=2	7∩3∩2=1	30∩23∩7=5
overlapped-by	11∩5∩4=2	59∩52∩42=20	3∩3∩2=0	0∩0∩1=0	73∩60∩49=22
finishes	2∩8∩1=0	10∩1∩11=0	5∩8∩5=1	1∩0∩0=0	18∩17∩17=1
during	449∩424∩650=217	105∩100∩113=62	206∩139∩225=67	112∩86∩134=43	872∩749∩1122=389
started-by	1∩0∩0=0	9∩2∩8=0	3∩14∩6=2	0∩3∩0=0	13∩19∩14=2
equal	1∩17∩0=0	37∩70∩51=19	263∩412∩307=154	62∩140∩90=29	363∩639∩448=202
starts	2∩0∩0=0	30∩9∩14=2	6∩16∩2=0	0∩1∩1=0	38∩26∩17=2
contains	164∩85∩144=63	830∩853∩868=671	299∩292∩344=117	148∩152∩188=64	1441∩1382∩1544=915
finished-by	0∩0∩0=0	3∩3∩0=0	6∩7∩6=0	1∩3∩0=0	10∩13∩6=0
overlaps	2∩2∩4=1	75∩84∩70=32	6∩27∩5=0	1∩4∩3=0	84∩117∩82=33
meets	1∩13∩0=0	25∩26∩2=2	88∩88∩32=22	9∩15∩0=0	123∩142∩34=24
before	739∩767∩746=572	389∩360∩383=288	1058∩994∩1098=713	418∩436=422=294	2604∩2557∩2649=1867
is_included	0∩0∩0=0	0∩0∩0=0	19∩2∩6=1	6∩0∩1=0	25∩2∩7=1
identity	0∩0∩0=0	0∩0∩1=0	11∩7∩24=2	16∩5∩15=2	27∩12∩40=4
includes	0∩0∩0=0	0∩0∩0=0	27∩10∩2=1	18∩2∩0=0	45∩12∩2=1
vague	115∩191∩157=38	212∩177∩191=100	327∩309∩265=128	154∩111∩111=48	808∩788∩724=314

Annotator A ∩ Annotator B ∩ Annotator C = Agreed count

Table 7 shows agreement rates by relation type across the three evaluation schemata. We define the schemata as follows. ‘Label 13+3+1’ is the most fine-grained evaluation schema; in it, all 13+3+1 relations are discriminated. ‘Label 13+1’ is a schema without event-subevent discrimination, in which ‘is\_included,’ ‘identity,’ and ‘includes’ are regarded in the same way as ‘during,’ ‘equals,’ and ‘contains,’ respectively. ‘Label 5+1’ is a TempEval-like schema, in which 13+3+1 relations are generalised into 5+1 relations: ‘BEFORE,’ ‘BEFORE-OR-OVERLAP,’ ‘OVERLAP,’ ‘OVERLAP-OR-AFTER,’ ‘AFTER,’ and ‘VAGUE’.

The agreement rate across all relations is 65.3% (Cohen’s kappa 0.733) using the most fine-grained evaluation schema (Label 13+3+1). We perform <TLINK> annotations on fixed

relation pairs for four types. TimeBank 1.2 jointly performs <TLINK> annotations without fixing the relation pairs. In this method, the <TLINK> relation agreement rate is 77% and the relation pair agreement 55%. We believe that the BCCWJ-TimeBank <TLINK> relation agreement rate is in no way inferior to that of TimeBank 1.2. Among the four relation types, the agreement rate of ‘DCT’ is the highest and that of ‘T2E’ is the second highest. The relation between a temporal information expression and an event instance is easier to determine than the relation between two event instances. This is because the interval of the temporal information expression is more easily defined on the timeline than the interval of the event instance is. Under the relaxed relation evaluation schema, the agreement rates of ‘E2E’ and ‘MATRIX’ increase. This means that, while interval boundary matching in these event instances is hard for the annotators to agree upon, interval anteroposterior relations can be agreed upon.

**Table 7. <TLINK>: Annotation agreement by relation type across three evaluation schemata.**

Relation types	DCT	T2E	E2E	MATRIX	ALL
Total Count	3839	2188	2972	1245	10244
Label 13+3+1	0.743(2854)	0.691(1513)	0.552(1642)	0.545(679)	0.653(6688)
Label 13+1	0.743(2854)	0.691(1513)	0.561(1667)	0.560(697)	0.657(6731)
Label 5+1	0.747(2873)	0.734(1605)	0.627(1862)	0.623(776)	0.695(7116)

Agreement rate(Agreed count)

Finally, Table 8 shows agreement by two entity types: DCT and TIMEX of <EVENT>@class. Relations with STATE tend to show low agreement rates, and relations between DCT/TIMEX and STATE are lower than relations between DCT/TIMEX and other event instances. This is because recognition of the time interval boundaries of state expressions is difficult for the annotators. In event instances with event arguments, relations with REPORTING and I\_ACTION tend to show higher than average agreement rates.

## Temporal and Event Information Annotation on Japanese Text

**Table 8.** <TLINK>: Annotation agreement by {DCT, <TIMEX3>, <EVENT>@class} × <EVENT>@class.

	DCT	TIMEX	OCC	REP	PER	ASP	I_A	I_S	STA	ALL
OCCURRENCE	0.739	0.702	0.551	0.625	0.286	0.718	0.559	0.592	0.422	0.656
Abbr. OCC	(2352)	(1320)	(1602)	(104)	(7)	(39)	(494)	(130)	(102)	(6159)
REPORTING	0.881	0.697	0.663	0.222	1.000	0.667	0.519	0.368	0.500	0.694
Abbr. REP	(126)	(66)	(95)	(9)	(2)	(3)	(52)	(19)	(12)	(385)
PERCEPTION	0.815	0.700	0.444	NaN	0.000	NaN	0.500	1.000	0.000	0.646
Abbr. PER	(27)	(10)	(18)	(0)	(1)	(0)	(6)	(1)	(1)	(65)
ASPECTUAL	0.714	0.615	0.545	1.000	0.000	0.000	0.643	0.000	0.000	0.627
Abbr. ASP	(63)	(52)	(44)	(6)	(2)	(2)	(14)	(1)	(1)	(185)
I_ACTION	0.808	0.720	0.576	0.690	0.667	0.765	0.631	0.527	0.333	0.698
Abbr. I_A	(880)	(567)	(491)	(29)	(6)	(17)	(309)	(55)	(51)	(2407)
I_STATE	0.651	0.686	0.490	0.250	0.750	0.429	0.545	0.875	0.333	0.594
Abbr. I_S	(195)	(86)	(145)	(4)	(4)	(7)	(55)	(16)	(15)	(527)
STATE	0.492	0.398	0.356	0.600	1.000	0.444	0.431	0.333	0.238	0.424
Abbr. STA	(181)	(83)	(118)	(5)	(3)	(9)	(51)	(9)	(21)	(481)
ALL	0.743	0.691	0.548	0.618	0.560	0.649	0.573	0.562	0.374	0.653
	(3839)	(2188)	(2524)	(157)	(25)	(77)	(984)	(233)	(203)	(10244)

Agreement rate

(Agreed count)

## 6. Conclusion

This paper has presented temporal information annotation on BCCWJ. This is the first corpus-based study on Japanese temporal information annotation with the ISO-TimeML standard. We adapted the temporal information annotation specification of the original TimeML and ISO-TimeML to the Japanese language in several layers: <TIMEX3>, <EVENT>, <MAKEINSTANCE>, and <TLINK>. In addition, we introduced two annotation paradigms for linguistic research on Japanese temporal information: the MAMA-cycle annotation with a pair-programming method on <TIMEX3> and the annotation as a cognitive experiment on <TLINK>. Finally, we evaluated the cognitive process in human temporal information processing. The achieved temporal ordering agreement rates were 65.3%.

In future research, we will continue to investigate machine-learning-based temporal ordering estimation. In English temporal ordering, the tense and aspect information in <MAKEINSTANCE> are important features. In Japanese temporal ordering, however, the

morphologically overt information is ‘る(-ru)’ vs. ‘た(-ta)’ for non-past and past tense, and ‘る(-ru)’ vs. ‘ている(-teiru)’ for limited aspect. We will report the results of this temporal ordering estimation in future publications.

We also intend to take advantage of BCCWJ’s status as the first balanced large-scale shared corpus of Japanese by analysing our annotation in comparison to the syntactic and semantic annotations conducted on BCCWJ by several Japanese research institutes, as mentioned in Section 2.2. Since Japanese is a modality-rich language, the modality annotations by these other institutes will be important for temporal ordering.

Furthermore, we will continue to evaluate the cognitive process in human temporal information processing. In this study, the annotators were professionally trained. In our next study, we will use crowdsourcing with 200 experimental subjects for the temporal ordering annotation. We will compare the annotation results between trained annotators and crowdsourcing subjects and will evaluate any differences between the specialists and non-specialists to further our understanding of human temporal information processing.

## Reference

- Boguraev, B., & Kubota Ando, R. (2005). TimeML-compliant text analysis for temporal reasoning. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI-05)*, 997-1003.
- Cheng, Y., Asahara, M., & Matsumoto, Y. (2008a). Constructing a temporal relation tagged corpus of Chinese on dependency structure analysis. *Journal of Association for Computational Linguistics and Chinese Language Processing (CLCLP)*, 13(2), 171-196.
- Cheng, Y., Asahara, M., & Matsumoto, Y. (2008b). The effect of event type classification for temporal relation identification in Chinese text. *The International Journal of Computer Processing of Oriental Language*, 21(2), 151-167.
- Den, Y., Nakamura, J., Ogiso, T., & Ogura, H. (2008). A proper approach to Japanese morphological analysis: Dictionary, model, and evaluation. In *Proceedings of the 6th Language Resources and Evaluation Conference (LREC-2008)*, 1019-1024.
- Grishman, R., & Sundheim, B. (1996). Message understanding conference-6: A brief history. In *Proceedings of Volume 1: The 16th International Conference on Computational Linguistics (COLING-96)*, 466-471.
- Im, S., You, H., Jang, H., Nam, S., & Shin, H. (2009). KTimeML: Specification of temporal and event expressions in Korean text. In *Proceedings of the 7th Workshop on Asian Language Resources (ACL-IJCNLP 2009)*, 115-122.
- Kudo, M. (1995). Asupekuto, Tensu-Taikei to Tekusuto --- Gendai Nihongo no Jikan-no Hyougen (in Japanese). ‘Aspect, tense and text, time expression in contemporary Japanese’, Hituzi Syobo.

*Temporal and Event Information Annotation on Japanese Text*

- Kudo, M. (2004). Nihongo-no Asupekuto, Tensu, Muudo-Taikai Hyouujun-go Kenkyu-wo Koete (in Japanese). ‘Aspect, tense, mood of Japanese, beyond the standard Japanese’, Hituzi Syobo.
- Li, H., Strötgen, J., Zell, J., & Gertz, M. (2014). Chinese temporal tagging with HeidelTime. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL-2014)*, 133-177.
- Maekawa, K. (2008). Balanced corpus of contemporary written Japanese. In *Proceedings of the 6th Workshop on Asian Language Resources (ALR)*, 101-102. Hyderabad, India. Association for Computational Linguistics.
- Mani, I. (2006). Machine learning of temporal relations. In *Proceedings of the 44th Annual Meeting of the Association for Computational Linguistics (ACL-2006)*, 753-760. Sydney, Australia. Association for Computational Linguistics.
- Nakamura, C. (2001). Nihongo-no Jikan Hyougen (in Japanese). ‘Time expressions in Japanese’, Kuroshio Shuppan.
- Pustejovsky, J., Castaño, J., Ingria, R., Saurí, R., Gaizauskas, R., Setzer, A., *et al.*, (2003a). TimeML: Robust specification of event and temporal expressions in text. In *Proceedings of the 5th International Workshop on Computational Semantics (IWCS-5)*, 337-353. Uppsala, Sweden. Association for Computational Linguistics.
- Pustejovsky, J., Hanks, P., Saurí, R., See, A., Gaizauskas, R., Setzer, A., *et al.*, (2003b). The TimeBank corpus. In *Proceedings of Corpus Linguistics 2003*, 647-656. Lancaster, UK. UCREL technical paper (number 16).
- Pustejovsky, J., Lee, K., Bunt, H., & Romary, L. (2010). ISO-TimeML: An international standard for semantic annotation. In *Proceedings of the 6th Language Resources and Evaluation Conference (LREC-2010)*, 394-397. Valletta, Malta. European Language Resources Association.
- Pustejovsky, J., & Stubs, A. (2012). *Natural Language Annotation for Machine Learning*, O’Reilly.
- Sekine, S., & Isahara, H. (2000). IREX: IR and IE evaluation project in Japanese. In *The Proceedings of the Second International Conference on Language Resources and Evaluation (LREC 2000)*, 1475-1480. Athens, Greece. European Language Resources Association.
- Sekine, S., Sudo, K., & Nobata, C. (2002) Extended named entity hierarchy. In *The Proceedings of the Third International Conference on Language Resources and Evaluation (LREC 2002)*. 1818-1824. Las Palmas, Canary Islands, Spain. European Language Resources Association.
- UzZaman, N., Llorens, H., Derczynski, L., Allen, J., Verhagen, M., & Pustejovsky, J. (2013). Semeval-2013 task 1: Tempeval-3: Evaluating time expressions, events, and temporal relations. Second Joint Conference on Lexical and Computational Semantics (\*SEM), Volume 2: In *Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, 1-9. Atlanta, Georgia, USA, Association for Computational Linguistics.

- Verhagen, M., Gaizaukas, R., Schilder, F., Hepple, M., Katz, G., & Pustejovsky, J. (2007). Semeval-2007 task 15: Tempeval temporal relation identification. In *Proceedings of the 4th International Workshop on SemEval-2007*, 75-80. Prague, Czech Republic, Association for Computational Linguistics.
- Verhagen, M., Saurí, R., Caselli, T., & Pustejovsky, J. (2010). SemEval-2010 Task 13: TempEval-2. In *Proceedings of the 5th International Workshop on Semantic Evaluations (SemEval-2010)*, 57-62. Stroudsburg, PA, USA, Association for Computational Linguistics.
- Warker, C., Strassel, S., Medero, J., & Maeda, K. (2006). ACE 2005 Multilingual Training Corpus. *Linguistic Data Consortium*, Philadelphia.
- Xue, N., Xia, F., Chiou, F-D., & Parmer, M. (2005). The Penn Chinese TreeBank: Phrase Structure Annotation of a Large Corpus. *Natural Language Engineering*, 11(2), 207-238.
- Xue, N., & Zhou, Y. (2010). Applying syntactic, semantic and discourse constraints in Chinese temporal annotation. In *Proceedings of the 23rd International Conference on Computational Linguistics (COLING-2010)*: Poster Volume, 1363-1372.